Fermilab

TM-988
0402.000

A STUDY OF THE WIDTH OF NONLINEAR RESONANCE

S. Ohnuma

September 15, 1980

I. Introduction

The word "study"* which appears in the title of this report may be regarded as a reflection of the fact that there are more questions raised here than answers given. This work is a byproduct of my frustrating effort to understand several reports on the subject of nonlinear resonances. More specifically, the central question raised in this report is on the precise meaning of "width" of an isolated nonlinear resonance when the concept of "width" and its specific values are used to discuss the stability of beams in particle accelerators and storage rings. In raising this question, I am well aware of a definite possibility of being regarded as unnecessarily fussy. Many people would undoubtedly agree with Alex Chao who says

> "One should not, however, take the definitions of resonance width too seriously since they are not rigorously defined quantities and serve only as order of magnitude concepts."[1]

I do not necessarily agree with this statement but it certainly is not important to have a difference of, say, a factor two or π as long as one is consistent in the use of resonance width and, far more important, as long as the physical meaning of the definition is clear and reasonable. I must confess that my frustration is caused at least partially by the carelessness I see in some reports.

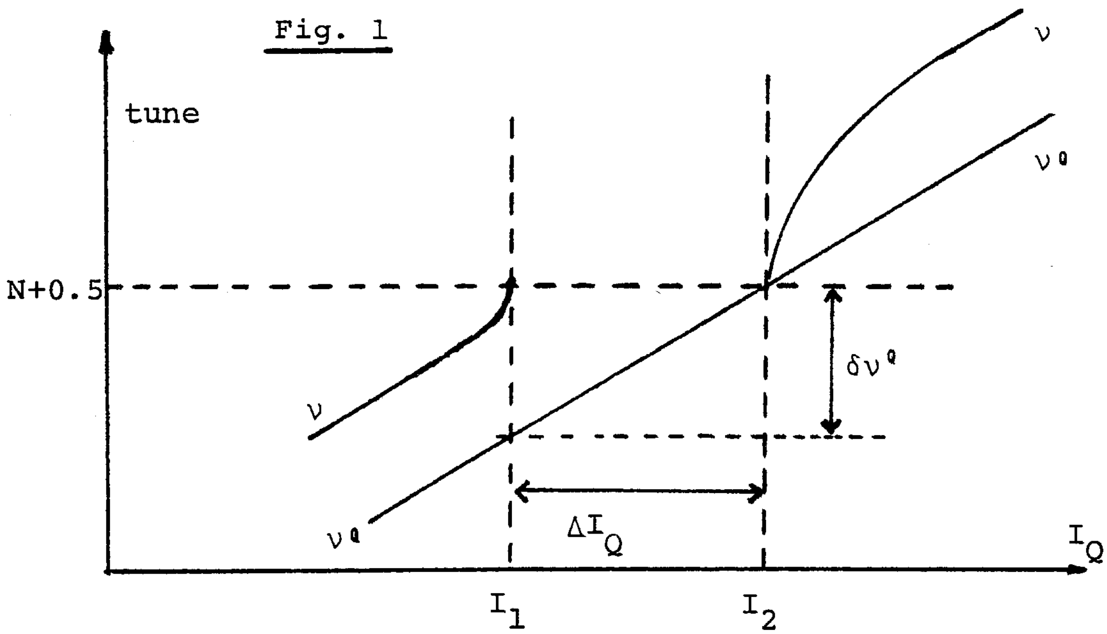
---

* "a careful examination or analysis of a phenomenon, development, or question;" - Webster's Dictionary -

II. Resonance Width and Stable Region in Phase Space

We start with the case for which there should be very little ambiguities in principle. Consider a linear ring, that is, a ring with dipole and quadrupole fields only. Furthermore, assume that the ring is made of two identical superperiods. When the betatron tune ν^o of this machine is near an integer plus 0.5, it varies almost linearly as a function of the quadrupole current I_Q for a fixed beam momentum. This is schematically shown in Fig. 1.

Fig. 1



The motion is stable regardless of the magnitude of beam emittance as long as the beam pipe is large enough to take care of the (finite) beam size everywhere around the ring. Now introduce a quadrupole somewhere in the ring and excite it at a certain level. The tune ν as a function of the quadrupole current I_Q is then quite different from the unperturbed case, again as shown in Fig. 1. As the current approaches a certain value I_1, the tune increases toward (N+0.5) and the beam becomes unstable. The maximum beam size is theoretically infinite regardless of the size of emittance. When the current I_Q reaches I_2, the stability of the beam is recovered and the tune varies again more or less linearly with I_Q. The "stopband width" of linear resonance $2\nu = 2N + 1$ is then $\delta\nu^{\circ}$ as shown in Fig. 1. Operationally, it may be less ambiguous to define the width in terms of the gap in I_Q, that is, ΔI_Q. However, even the perturbed tune ν is quite linear in I_Q and there is no difficulty in translating from ΔI_Q to $\delta\nu^{\circ}$. The most important feature of linear resonances when compared to the nonlinear ones is that the entire phase space is either stable or unstable and the beam emittance does not enter in the definition of the resonance width.

Nonlinear resonances are not much more complicated than the linear ones when only one degree of freedom is involved and the resonance under consideration is isolated. The two-dimensional phase space is divided into two parts, the central area in which the motion is confined and the outer area where the amplitude of motion may or may not be finite. Fig. 2 is an example of the fourth-order resonance, 4ν = an integer + ε, where coordinates are suitably normalized such that the beam is represented by a circle when the nonlinear field is turned off.

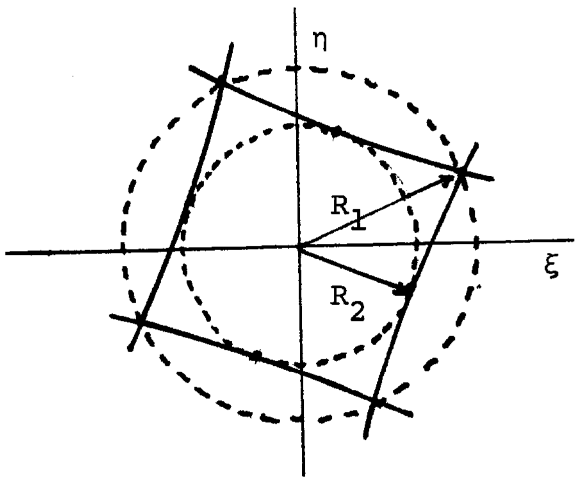


Fig. 2

For $\varepsilon = 0$, exactly on the resonance, there is no central stable area. The "width" of resonance is defined as the range of ν for which the central area is less than the beam emittance so that, if the tune is within this range, part of the beam may become unstable. There are of course variations in the definition; for example, one may define the width as the range of ε instead of the range of ν. In place of the area of the central stable region, one may prefer to use either

$\pi R_1^{\ 2}$ or $\pi R_2^{\ 2}$, the former giving a smaller value and the latter a larger value as the width compared to the standard difinition given above. These variations are not important as long as one is consistent in the definition when one tries to compare two or more cases. The size of the resonance width now clearly depends on the beam emittance E and this dependence becomes stronger as the order of resonace n for $n\nu = N$ increases.*

The picture becomes rather comlex when two degrees of freedom are involved in nonlinear resonances. In what follows, for the sake of simplicity of treatment, the so-called "shear terms" or "phase-independent terms" in the Hamiltonian are ignored. These terms arise from magnetic fields of octupole, 12-pole, 16-pole, etc. averaged around the ring. One may argue that, in a real machine, this is not too unrealistic since one should try to have a correction system that will eliminate at least the average octupole field. On the other hand, by ignoring the shear terms, one loses the most important property of nonlinear oscillations, the dependence of tunes on the oscillation amplitudes in the lowest-order approximation. This causes difficulties in trying to understand multi-resonance effects (see the next section). Another simplification introduced here is to limit the discussion to sum resonances only, that is, resonances of the form

$$m \cdot \nu_1 + n \cdot \nu_2 = k \qquad (\, m \leqslant n)$$

with both m and n positive, $m + n > 2$. It is well-known that, for difference resonances $m \cdot n < 0$, oscillation amplitudes are always finite. Consequently, any definition of resonance width would necessarily require an arbitrary parameter such as the maximum allowed amount of energy transfered from one direction to the other.

There are several reports[2-9] on the subject of the width of nonlinear coupled resonances. The one by Sturrock[2] is somewhat unorthodox in its formalism but I have found it to be extremely valuable. A detailed discussion on coupling resonance is given for $\nu_1 + 2\nu_2 = k$

* The width is proportional to $\sqrt{E}^{\,n-2}$.

but the method can be applied in principle to any other resonances. (It is hard for me to understand why this report and the one by Lysenko[3] are often ignored by others in the discussion of nonlinear resonances, especially in view of the fact that both are published papers and not informal laboratory reports.) A statement made by Sturrock (page 178 of ref. 2) is probably more responsible than anything else for arousing my interest in the subject:

> "The most surprising feature of the stability diagram of Fig. 28 is that there are points of arbitrarily small amplitudes u, v, which lie outside the stable region."

This is equivalent to saying that, for a beam of very small emittances, the width of resonance can be very large if the ratio of two emittances satisfies a certain condition. Particles in the central area of a beam can then become more unstable compared to those in the outer area which is hard to understand. I am not sure if this statement has been questioned by someone; Lysenko[3] has given the correct picture of the resonance $\nu_1 + n\nu_2 = k$ (Fig. 4 of ref. 3) but he does not explicitly point out the error made by Sturrock. Recently, Guignard at CERN has worked out a quite general treatment of all sum and difference resonances in a three-dimensional (compared to the usual two-dimensional) magnetic field.[9] According to this work, "a bandwidth inside which the amplitudes can become infinite" can be written as*

$$\Delta e \propto E_1^{(m-2)/2} \, E_2^{(n-2)/2} \, [m^2 E_2 + n^2 E_1] \qquad (1)$$

for resonance $m \cdot \nu_1 + n \cdot \nu_2 = k$ where E_1 and E_2 are the <u>initial</u> emittances of the beam in two directions. This expression seems to be widely used at CERN as <u>the</u> bandwidth, especially in works related to the beam storage problems. The more detailed form of width is not really important for the present purpose except for the fact that it is linearly dependent on the strength of the field that is responsible

---

* Eq. (2.6.2) of CERN-ISR-MA/75-35 or Eq. (6.17) of CERN 78-11, both of ref. 9.

for the resonance and that, by applying this expression to uncoupled resonances, one can easily see that it corresponds to the choice of $\pi R_1^{\ 2}$ (see Fig. 2) being equal to the beam emittance. The difinition of the bandwidth should therefore be "inside which the amplitudes always become infinite regardless of the initial phases". Again this is simply a matter of one's preference and we should only be aware of the physical meaning of this definition in applying the formula for some practical cases. A more serious defect of this expression is that, for either m (or n) = 1, the bandwidth increases indefinitely as E_1 approaches zero while E_2 is fixed. This is schematically shown in Fig. 3.

This unphysical property of Eq. (1) is related to the erroneous statement made by Sturrock (see p. 5) for resonance $\nu_1 + 2\nu_2 = k$. This point has been examined recently[6] and the following is a summary of my present understanding. Related material can also be found in ref. 4 for resonances $2\nu_1 + 2\nu_2 = k$ and $\nu_1 + 3\nu_2 = k$.

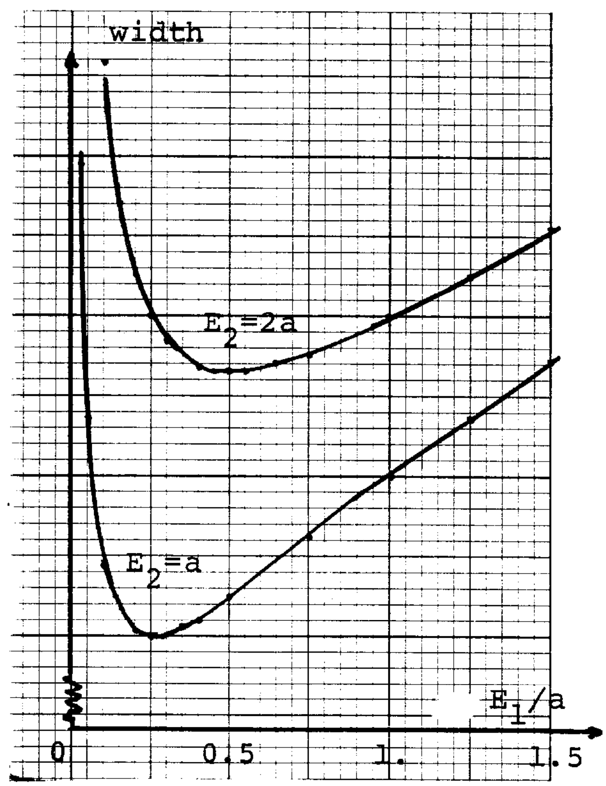


Fig. 3

In terms of the action-angle variables (I, a), the Hamiltonian for an isolated resonance $m \cdot \nu_1 + n \cdot \nu_2 = k + \varepsilon$ is, in the absence of shear terms,

$$H(I_1, a_1, I_2, a_2; \theta)$$

$$= (\varepsilon_1/2)(2I_1) + (\varepsilon_2/2)(2I_2) + D \cdot \cos(\phi)(2I_1)^{m/2}(2I_2)^{n/2} \qquad (2)$$

where $\phi \equiv m \cdot a_1 + n \cdot a_2 + \delta$. The amplitude D and the phase δ of the driving term can be expressed in terms of the machine parameters and the parameters of the nonlinear field which is driving the resonance. With the nearest point (ν_{1o}, ν_{2o}) on the resonance line, we have $m \cdot \nu_{1o} + n \cdot \nu_{2o} = k$, $\nu_1 \equiv \nu_{1o} + \varepsilon_1$ and $\nu_2 \equiv \nu_{2o} + \varepsilon_2$. Since H does not depend on the independent variable θ explicitly, it is an invariant of the motion. In trying to define the width of resonance with two degrees of freedom, one may follow an idea used for one degree of freedom. There one finds unstable fixed points and equate the distance of these points from the origin in phase space to the beam emittance. This is equivalent to saying that the beam emittance is equal to $\pi R_1^{\,2}$ in Fig. 2. Analogous to the concept of unstable fixed points in the two-dimensional phase space, one defines "fixed lines"* in the four-dimensional space from the three conditions,

$$dI_1/d\theta = dI_2/d\theta = d\phi/d\theta = 0.$$

Since the beam emittance E_i and the action variable I_i are related,

$$E_1 = \pi(2I_1), \qquad E_2 = \pi(2I_2),$$

these three conditions for fixed lines lead to Eq. (1), Guignard's formula for the bandwidth, in a straightforward manner. To me, this is somehow more reasonable way to derive the expression than the one given by Guignard (pp. 32 - 37, CERN 78-11). In the case of two-dimensional phase space, one can see pictorially that the use of unstable fixed points is physically reasonable in defining the width of resonance. It is not so obvious in four-dimensional phase space; one must have some other independent assurances that it is physically meaningful to relate the fixed lines with the beam emittances to define the bandwidth. When this is done, as in ref. 6, one finds that resonances of the form $\nu_1 + n \cdot \nu_2 = k$ are different from other resonances in their behavior in phase space and this difference can

* The term "fixed lines" was used by A. Ruggiero in his report[8] on beam-beam interaction.

account for the paradoxical results obtained by Sturrock and by Guignard. In this connection, the work by Ruggiero on two-dimensional beam-beam interaction[8] is quite intriguing. Of course his main concern was not a general treatment of nonlinear resonances but rather the special characteristics of beam-beam interaction and the stochasticity limit of stability (see the next section) arising from this interaction. In schematically showing the pictures in phase space (Figs. 1a and 1b on p. 12 of FN-258), he must have been aware of the special feature of resonances $\nu_1 + n\cdot\nu_2 = k$. On page 11 of FN-258, there is a statement

"The other case with $|q_o| \leqslant 1$ will be considered later."

Unfortunately, the promised consideration does not materialize in the report. It should also be mentioned that the condition $|q_o| \leqslant 1$ is not completely equivalent to the condition m (or n) = 1 in $m\cdot\nu_1 + n\cdot\nu_2 = k$.

From equations of motion for I_1 and I_2 with the Hamiltonian (2) on p. 6, one sees that

$$C \equiv (2I_1)/m - (2I_2)/n \tag{3}$$

is another invariant of the motion. From H and C, one can construct, with a suitable normalization of variables,[6] two invariants of the form

$$\begin{aligned} \lambda &= u^2 + u^m v^n w, \\ \mu &= v^2 + u^m v^n w. \end{aligned} \tag{4}$$

Two amplitudes u and v are proportional to $\sqrt{I_1}$ and $\sqrt{I_2}$, respectively. For physically meaningful solutions, one must have

$$u \geqslant 0, \quad v \geqslant 0, \quad \text{and} \quad |w| \leqslant 1.$$

All nonlinear resonances ($m + n \geqslant 3$, $m \leqslant n$) can be classified to two groups, one with $m \neq 1$ and the other with $m = 1$.

1) <u>$m \neq 1$</u> . The lowest-order resonance of this group is $2 \cdot \nu_1 + 2 \cdot \nu_2 = k$. For any resonance of this group, the function $w(u)$ has only one minimum value (and no maximum value). If the minimum value $w(u_M)$ is larger than -1, the motion is unstable since u can take any value within $|w| \leqslant 1$. For $w(u_M) < -1$, a stable motion is possible and the limiting case, shown in Fig. 4a, corresponds to $w(u_M) = -1$.

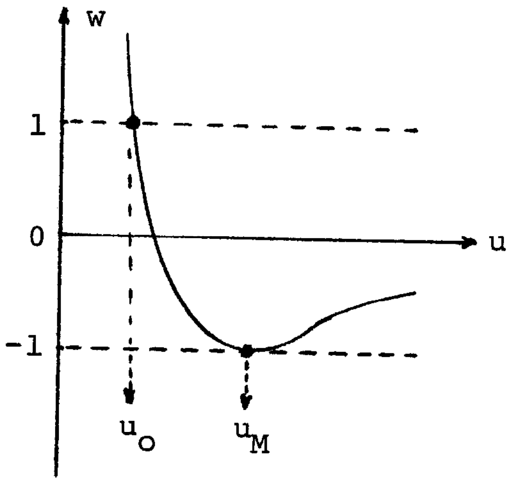


Fig. 4a

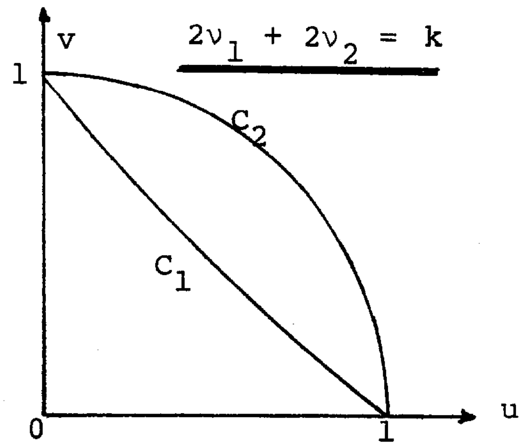


Fig. 4b

When the initial amplitude is less than or equal to u_o in Fig. 4a, the motion is stable regardless of initial values of two phases a_1 and a_2. If, on the other hand, the initial value of u exceeds u_M, the motion is unstable independent of phase. In between, the initial phase (or more specifically, the combination $m \cdot a_1 + n \cdot a_2$) will determine whether a stable motion is possible or not. Of course the value of the other amplitude v cannot be arbitrary in order for the motion to be stable and the limiting boundary in (u, v) space looks like the one shown in Fig. 4b for $2 \cdot \nu_1 + 2 \cdot \nu_2 = k$. (The picture is identical to Fig. 1b of FN-258, ref. 8.) Here the curve C_1 is the limiting relation for (u_o, v_o) and C_2 for (u_M, v_M). From the nature of (u_M, v_M), it is obvious that the curve C_2 is identical to Eq. (1), the bandwidth of Guignard. In terms of (u_M, v_M), the relation is

$$(m/2)\cdot u_M^{\,m-2}\, v_M^{\,n} \quad + \quad (n/2)\cdot u_M^{\,m}\, v_M^{\,n-2} = 1 \qquad (5)$$

which takes a particularly simple form

$$u_M^2 + v_M^2 = 1$$

for $2\cdot\nu_1 + 2\cdot\nu_2 = k$. In general, it is not possible to write down an algebraic expression for the curve C_1 (u_o, v_o) although even this can be done[4] for the lowest-order case in Fig. 4b. The relation is

$$v_o^{\,2} = (1 + 3u_o^{\,2}) - \sqrt{(1 + 3u_o^{\,2})^2 - (1 - u_o^{\,2})^2}.$$

2) $\underline{m = 1}$. The lowest-order resonance of this group is $\nu_1 + 2\cdot\nu_2 = k$. In addition to the one shown in Fig. 4a, there is another type of limiting case with one minimum and one maximum points. This is shown in Fig. 5a. It should be noted here that, in (λ, μ) space, the situation depicted in Fig. 4a corresponds to $\lambda \geqslant 0$ and $\mu > 0$ while the one in Fig. 5a is for $\lambda < 0$ and $\mu > 0$.

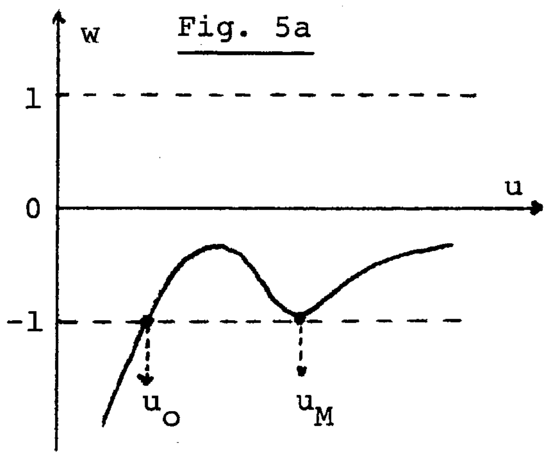


Fig. 5a

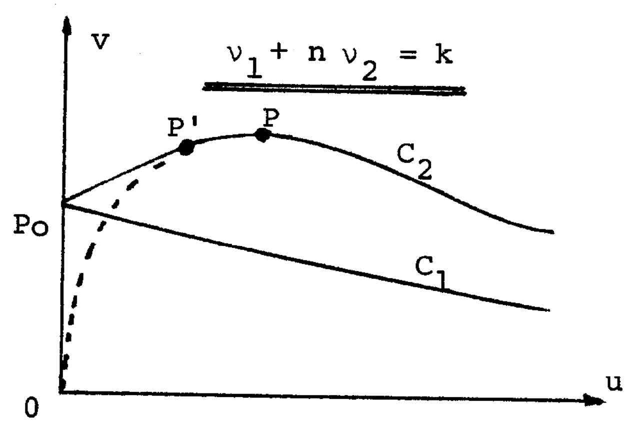


Fig. 5b

If one considers the case given in Fig. 4a, (u_M, v_M) still satisfy the relation (5) and, with $m = 1$, the curve C_2 must pass the origin $u = v = 0$. This is shown in Fig. 5b where C_2 includes the dotted curve past point P' to the origin. Points above this curve are supposedly unstable. They can be as close to the origin as one wishes and the indefinitely increasing bandwidth is a direct consequence of this behavior. Another way of seeing the same feature is to insist that λ must be positive (as Sturrock did). Since this condition must be valid for any value of w including $w = -1$, positive λ for $m = 1$ means $u > v^n$ which again eliminates points near the origin from the stable area. By including the case in Fig. 5a, one can show that the true boundary C_2 is the solid curve in Fig. 5b. Any motion that starts from the curve C_1 cannot go beyond the curve C_2 and is limited to the right of point P. Points between P and P' can be reached if the motion starts from points between P' and P_o.

Once the boundaries C_1 and C_2 are found, either algebraically or numerically, the bandwidth can be evaluated by putting the beam within C_1 ("πR_2^2" version in Fig. 2) or C_2 ("πR_1^2" version in Fig. 2). Depending on the particle distribution in (u, v) space,* one may take

$$u \leqslant A \quad \text{and} \quad v \leqslant A/\alpha$$

as in Fig. 6a or

$$u^2 + \alpha^2 \cdot v^2 \leqslant A^2$$

as in Fig. 6b with α and A specifiying the beam size.

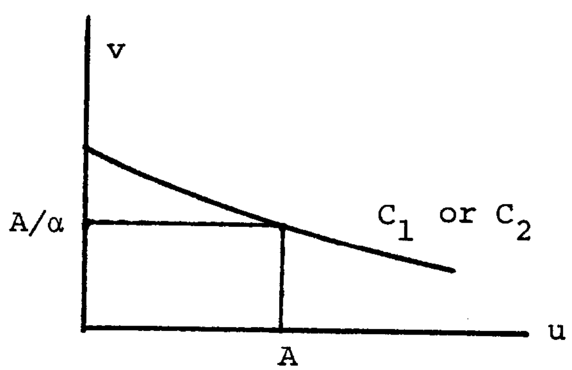


Fig. 6a

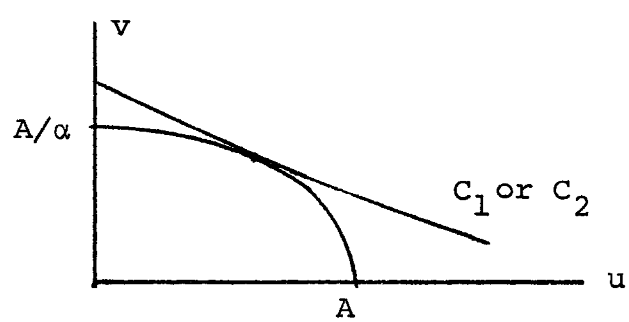


Fig. 6b

*Distributions are assumed to be uniform in phases.

For distributions with a long tail, e. g., Gaussian, one must of course specify the fraction of the beam defined by α and A. For some cases, it may be more reasonable to assume that the resonance is adiabatically turned on and the bandwidth should be evaluated by equating the phase space volume bounded by $u = 0$, $v = 0$ and the curve C_1. Strictly speaking, this is an underestimate of the allowed volume (which leads to an overestimate of the bandwidth) since the beam may occupy the region between C_1 and C_2 for some phase values. The four-dimensional volume is

$$
\begin{aligned}
V &\equiv \iiiint dx\, dx'\, dy\, dy' = \iint dE_1\, dE_2 \\
&= E_o^2/\gamma \qquad\qquad \text{if } E_1 \leqslant E_o \text{ and } E_2 \leqslant E_o/\gamma \\
&= E_o^2/(2\gamma) \qquad\quad \text{if } E_1 + \gamma E_2 \leqslant E_o \,.
\end{aligned}
$$

The volume is proportional to the integral

$$\iint d(u^2)\cdot d(v^2) \tag{6}$$

where the integration is over the region bounded by $u = o$, $v = 0$ and C_1. I have found out that this can be done analytically for $\nu_1 + 2\nu_2$ and $2\nu_1 + 2\nu_2$. The resulting bandwidths agree reasonably well with the estimate given by Collins and Edwards[7] (see the table on p. 18 of ref. 7). For other resonances considered by them, that is, $\nu_1 + 3\nu_2$, $\nu_1 + 4\nu_2$ and $2\nu_1 + 3\nu_2$, I must confess I do not understand how they obtained the results. To begin with, it is impossible for me to find an algebraic expression for the curve C_1. I may resort to some numerical integration; it is quite possible that results given in ref. 7 are based on numerical evaluations. The trouble I get into is that I cannot even prove whether the integral (6) converges or not for $\nu_1 + 3\nu_2$ and $\nu_1 + 4\nu_2$. For $2\nu_1 + 3\nu_2$, it is possible to show that the integral is at least finite since the same integral over the region bounded by C_2 (instead of C_1) is finite. It should be noted again that the curve C_2 (for $m = 1$, beyond the point P' in Fig. 5b) is given algebraically by Eq. (5) and the integral (6) with C_2 as a boundary is relatively simple. Although this integral definitely diverges for $\nu_1 + 3\nu_2$ and $\nu_1 + 4\nu_2$, it may still converge if the boundary is C_1 instead of C_2. The only choice I have is either to stick with nonadiabatic definitions (Figs. 6a and 6b) or to use the results given in ref. 7 without really understanding how they came about. Considering the reputation of the authors, one might even argue that the latter choice is not at all unreasonable.

III. Resonance Width for the Overlap Criterion of Chirikov

According to an early report by Chirikov, Keil and Sessler,[12] the famous overlap criterion of Chirikov first appeared in early 60's. Since then, there have been numerous reports on the subject but the best and presumably the most authoritative is the one written by Chirikov recently.[10] For a quick explanation of the criterion, one may look up one of the following:

1. p. 5 - p. 9, ref. 11
2. p. 23, ref. 8
3. p. 9 (ISR-TH/72-25), ref. 13
4. p. 4 - p. 5, ref. 12

When nonlinear motions of a periodic system are plotted in phase space, there are in general three types of orbits as shown in Fig. 7 (see p. 14) which is taken from ref. 16. Two dots P and P' represent periodic orbits with rational tune values. Two continuous lines A and B are the examples of KAM surfaces. Since these surfaces are "impenetrable", the motion of particles bounded by the KAM surfaces is stable, that is, the oscillation amplitude is always finite. However, the phase motion in this region is seemingly random and the region is therefore called "stochastic layers". As one can see in Fig. 7, stochastic layers are prominent near separatrices. Chirikov calls stochastic layers an "embryo" of instabilities since instabilities generally develop from these layers. One of the fundamental problems in nonlinear dynamics is to find, for a given system, the limiting strength of non-

J. M. Greene, PPPL-1489 (Princeton Plasma Laboratory), Nov. '79

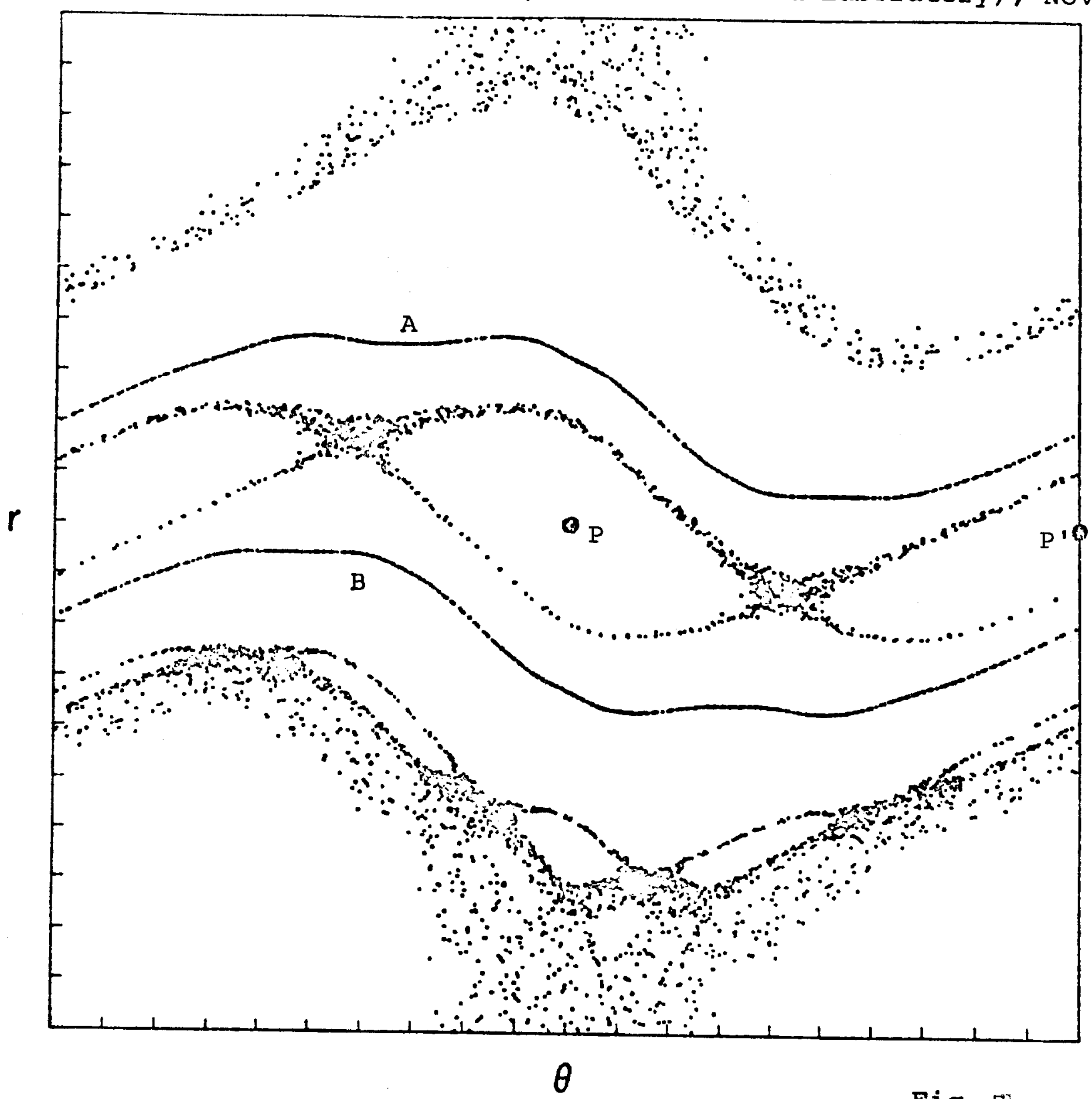


Fig. 7

liner field above which no KAM surfaces exist and the entire phase space becomes stochastic. The only theoretical guidance we have for this is the overlap criterion of Chirikov. Very simply stated, it conjectures that the motion becomes totally stochastic when the area

covered by all resonances in a square region in the tune diagram*
becomes unity. The total area is obtained by summing all resonance widths.

One can raise many questions regarding this criterion, especially on how to use it in real problems. Chirikov himself admits[10]: "The only excuse the overlap criterion may offer is that today any other theory is also incapable to answer the question ..." At the same time, he feels that it is "substantiated by plausible (at least, for a physicist!) considerations." It is undoubtedly "an order-of-magnitude estimate", again Chirikov's admission. One may just leave it at that except for the fact that many accelerator physicists have tried to find quantitatively the limiting strength of beam-beam interaction in storage rings based solely on this criterion. This may be partially justified; some numerical studies made for a few examples have shown that the criterion is a surprisingly good one even in quantitative sense. A more comprehensive discussion on the overlap criterion and its proper use is clearly beyond the intended scope of this report. I might just mention in passing that its popularity among accelerator physicists does not seem to be as much as it used to be only a year or so ago. In spite of inherent difficulties in any quantitative use of the criterion, and there are many, I cannot agree with Alex Chao (see p. 1) when he advises us not to take the definition of resonance width too seriously. Let me emphasize again that I am not fussing over a factor of two or π. What I do care strongly is that, when a value of resonance width is used for any purpose, the definition of width should be physically meaningful and everyone should be able to understand clearly what it really is. It is very frustrating for me to encounter several pages of complex mathematical expressions and numbers before I understand clearly what is calculated. Regardless of how one defines the width in connection with the overlap criterion, it must be a quantity one can evaluate for isolated resonances. This is unfortunate but unavoidable; unfortunate as the criterion is needed most for cases in which many resonances overlap

* ν_1 from N to N+1, ν_2 from N to N+1.

so that the very concept of an isolated resonance loses its validity, unavoidable since the assumption of isolated resonances is essential in almost all cases for an analytical treatment of nonlinear resonances. Furthermore, the width is calculated in the lowest-order approximation. This means the width is always linearly proportional to the strength of resonance-driving field.

Let me begin by quoting a passage from Chirikov (p. 287 of ref. 10),

> "A plausible condition for the occurrence of the stochastic instability seems to be the approach of resonances down to the distance of the order of a resonance size. Such an approach was naturally called the <u>resonance overlap</u>. To be precise, the overlap of resonances begins when their separatrices touch each other."

After reading this (and nothing else in the report), one may very well be tempted to identify the width to be used for the overlap criterion with the one discussed in the previous section.[14] Judging from what he says elsewhere in the same paper, I am inclined to say that this is completely wrong, or more specifically, that this is not in agreement with the physical picture Chirikov must have had in mind when he proposed the criterion. Still, I must confess I lack the conviction which is necessary to declare that such an identification of the width is totally devoid of any redeeming feature. In what follows, I will try to summarize simply but clearly Chirikov's definition of the resonance width so that we can cmpare that with other definitions used by various accelerator physicists. There will be many direct quotations from Chirikov.[10] They are made whenever I feel my rephrasing may cause misunderstandings or when Chirikov's true intention may be subject to individual reader's interpretation.

Chirikov emphasizes one particular feature of nonlinear oscillations as something very significant. It is the dependence of the oscillation frequency (tune) on the oscillation amplitude (energy). On p. 266, he says

> "A significant feature of the nonlinear resonance is the oscillation boundedness and smallness under a small perturbation as distinct from the linear resonance for which

> there is no such boundedness. The oscillations are bounded due to the dependence of their frequency on the energy. Such a dependence is, thus, an important property and the first sign of an oscillator nonlinearity."

On p. 268, under a title "What are nonlinear oscillations?",

> "We will expose the most important property of the non-liner oscillator, the so-called non-isochronicity, i.e., the dependence of the free oscillation period on the amplitude, or the energy."

As a measure of the degree of nonlinearity, he defines a parameter α in terms of the action variable I (p. 269),

$$\alpha \equiv \frac{I}{\nu}\frac{d\nu}{dI} \tag{7}$$

which is in general a function of I (but not of the angle variable). After excluding all questions related to the influence of dissipation, he states (p. 266)

> "Another essential restriction of the models in question is a limitation of the oscillation nonlinearity from below, that is we assume the nonlinearity to be not too small. .... As far as a theoretical analysis and estimates are concerned we have also to confine ourselves to the usual case of a small perturbation acting upon a system whose motion is known."

Notice the very important difference between "not-so-small" nonlinearity and a "small" perturbation. This is explained more explicitly on p. 270:

> "(A version of perturbation theory) is applicable, of course, only if there is a small parameter. According to this we assume that the Hamiltonian of our system may be divided into two parts:
>
> $$H(I, a) = H_0(I) + \varepsilon V(I, a) \qquad (2.\ 9)^*$$
>
> the first of which describes an "unperturbed" system and has an integral of motion I (the action). The main "property" of the unperturbed system is our complete knowledge of its motion. Our problem is, however, to study the motion of a "perturbed" system with the Hamiltonian H(I, a), and we assume the "perturbation"

* Chirikov uses the symbol θ as the angle variable which I changed to a here to avoid a possible confusion.

$\varepsilon V(I, a)$ to be small ($\varepsilon << 1$). A characteristic feature of the perturbation is the dependence of the latter on the phase a of the unperturbed motion. This very dependence leads to a change in the unperturbed action I."

(I apologize for these lengthy quotes; it is only hoped that, by citing all these passages, I am making it clear that I believe the underlying idea of Chirikov is essential to his thinking and to his overlap criterion.) Combining requirements for "not-too-small" nonlinearity, a small perturbation and isolated resonances altogether, Chirikov writes (p. 279)

$$\varepsilon << \alpha << \varepsilon^{-1}(\Delta\nu/\nu_o)^2$$

where ν_o is the tune on the resonance under consideration and $\Delta\nu$ is the distance from the actual tune ν to other resonance lines. The second of two inequalities here is clearly for the condition of isolated resonances. It is instructive to see how Chirikov derives his resonance width. We will consider a very simple uncoupled resonance $3\nu = k$ in the presence of sextupole and octupole fields. One can find the formalism on p. 278 - p. 279 of ref. 10 under the title "A universal description of a nonlinear resonance".*

In terms of action-angle variables (I, b), the Hamiltonian is

$$H(I, b; \theta) = H_o(I) + H_1(I, b), \tag{8}$$

$$H_o(I) = \nu_{oo} I + r_o(2I)^2,$$

$$H_1(I,b) = D\ (2I)^{3/2}\cos(3b - k\theta).$$

The angle variable b here is related to a in Eq. (2) by the relation

$$b = a + (k/3)\theta$$

and the phase δ of the driving term is assumed to be zero. The term $r_o(2I)^2$ in H_o is the shear term (phase-independent term) arising from the octupole field. In the absence of the driving term, $D = 0$, the

* While I was preparing this note, Fred Mills told me that the identical concept was introduced by Symon and Sessler many years ago and he himself worked out a number of relations although they have never been published.

tune is a function of the action variable,

$$\nu_o(I) = db/d\theta = \partial H_o/\partial I = \nu_{oo} + (4r_o)(2I) \tag{9}$$

where I is a constant of the motion, that is, the system is integrable. The parameter α, Eq. (7), to specity the nonlinearity is then

$$\alpha \equiv (I/\nu_o)(\partial\nu_o/\partial I) = (8r_o)(I/\nu_o) \tag{10}$$

A generating function is now introduced,

$$F(I, \phi;\theta) \equiv -(I - I_r)\cdot(\phi + k\theta)/3 \tag{11}$$

for transforming from (I, b) to (p, ϕ) with I_r to be specified later. Obviously,

$$b = -\partial F/\partial I = (\phi + k\theta)/3,$$

$$p = -\partial F/\partial\phi = (I - I_r)/3.$$

The new Hamiltonian is

$$\begin{aligned} K(p,\phi;\theta) &\equiv H + \partial F/\partial\theta = H - kp \\ &\approx H_o(I_r) + 3p[\nu_o(I_r) - k/3] + 36r_o p^2 + D(2I_r)^{3/2}\cos\phi \end{aligned}$$

In deriving this, it is assumed that $|p| \ll I_r$ and only the lowest-order term is retained in the small driving term. The choice of I_r is now obvious; it is made such that the term linear in p vanishes,

$$(2I_r) = -\frac{\nu_{oo} - k/3}{4r_o} \tag{12}$$

We assume that r_o is positive so that $\nu_{oo} < k/3$. The choice (12) is equivalent to saying that, when $D = 0$, the tune would be exactly on resonance at $I = I_r$. As a result, we have

$$K(p,\phi;\theta) \simeq p^2/(2M) + \varepsilon V(I_r)\cdot\cos(\phi) \quad [+\ \text{const.}] \tag{13}$$

where $M^{-1} = 72\ r_o = 9\ (\partial\nu_o/\partial I)$,

$$\varepsilon = (6D/k)(2I_r)^{1/2}, \qquad V(I_r) = \nu_o(I_r)\cdot I_r$$

The form (13) is familiar to accelerator builders as the Hamiltonian for a stationary RF bucket above transition.[15] The angular frequency of a small-amplitude phase oscillation is

$$\omega_\phi = \sqrt{(\varepsilon V/M)} = 3\nu_o(I_r)\cdot\sqrt{(\varepsilon\alpha)}. \tag{14}$$

The equation for the separatrix is

$$[p_s(\phi)]^2 = 4M\cdot\varepsilon V(I_r)\cdot\sin^2(\phi/2) \tag{15}$$

so that, on the separatrix,

$$\begin{aligned} I &\equiv I_r + 3\cdot p_s(\phi) \\ &= I_r \pm 6\sqrt{[M\varepsilon V(I_r)]}\cdot\sin\left(\frac{3b - k\theta}{2}\right). \end{aligned}$$

At this point, Chirikov says (p. 279, ref. 10) "The amplitude of phase oscillation $(\Delta I)_r$ may be called the resonance half-width (in I).",

$$(\Delta I)_r \equiv 6\sqrt{[M\varepsilon V(I_r)]} = 2\sqrt{(\varepsilon/\alpha)}\cdot I_r \tag{16}$$

"The resonance half-width in frequency is equal to", Chirikov continues,

$$(\Delta\nu)_r \equiv (\Delta I)_r\cdot(\partial\nu_o/\partial I) = (2/3)\omega_\phi = 2\nu_o(I_r)\sqrt{(\varepsilon\alpha)}. \tag{17}$$

It is quite clear from this definition, Eq. (17), that the resonance width Chirikov talks about has nothing to do with the one considered in the previous section , the one related to the beam emittance. This becomes clearer if one writes $(\Delta I)_r$ and $(\Delta\nu)_r$ in terms of r_o, D and $|\delta\nu| \equiv |\nu_{oo} - k/3|$:

$$(\Delta I)_r = (1/4)\cdot\sqrt{D}\, r_o^{-5/4} |\delta\nu|^{3/4} , \tag{18}$$

$$(\Delta\nu)_r = 2\cdot\sqrt{D}\, r_o^{-1/4} |\delta\nu|^{3/4} . \tag{19}$$

Note that the dependence of $(\Delta\nu)_r$ on the strength of shear term, r_o, is rather weak but this is not the case for $(\Delta I)_r$. The assumption $|p| \ll I_r$ needed to derive the Hamiltonian K on p. 19 is equivalent to the condition of "not-too-small" nonlinearity, $\varepsilon \ll \alpha$ (see p. 18)* and this can be expressed in the form

$$D^2 \ll r_o \cdot |\delta\nu| \tag{20}$$

It is important, I believe, that the unperturbed Hamiltonian H_o has shear terms ($r_o \neq 0$). Otherwise, it is not clear to me how one can define the resonance width in accordance with the spirit of Chirikov. Since the system is completely integrable, the existence of shear terms introduces no added mathematical difficulty. In other words, there is no excuse of dropping shear terms from the Hamiltonian unless they are truly absent in the system under consideration.

We are now ready to examine how various accelerator physicists interpreted the resonance width that should be used for the overlap criterion of Chirikov. The identification of the width with the one discussed in the previous section has been mentioned already. We start with the report by Chirikov, Keil and Sessler.[12] As one naturally expects from the fact that Chirikov himself is one of the authors, the definition of resonance width is identical to what has been discussed above. The importance of "not-too-small" nonlinearity parameter α (see p. 17) in the formalism is properly stressed and the overlap criterion is applied to four examples. However, there is a section (which consists of just thirteen lines) called "the third model" in which they attempt to deal with a case with very small non-

---

* The other inequality on p. 18, the one for an isolated resonance, is equivalent to " resonance half-width $(\Delta\nu)_r$ much less than the distance from the actual tune to resonances other than $3\nu = k$ ".

linearity. The dependence of ν on I arises only as a result of the small perturbation and α is of the order of ε, the small perturbation parameter. I simply do not understand the argument they present here; it is impossible to apply what they are saying (i.e., what I believe they are saying) to our example with the shear term $r_o = 0$. It is true that one can still define the effective tune ν_{eff} in the form

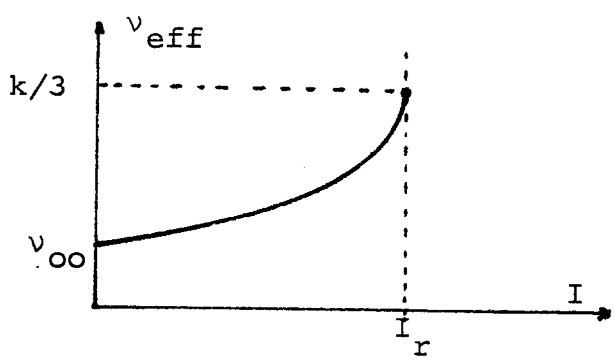
$$k/3 - \nu_{eff} = \int_0^{2\pi} \frac{d\phi}{|\delta\nu| - 3D\sqrt{2I}\cos(\phi)} \tag{21}$$

where $\delta\nu \equiv \nu_{oo} - k/3 < 0$. The integral can be done analytically if one assumes that D is sufficiently small and $I \simeq$ constant,

$$\nu_{eff}(I) \simeq k/3 - \sqrt{(\delta\nu)^2 - 9D^2(2I)} \,. \tag{22}$$

Note that the expression is exact at $I = 0$ (where $\nu_{eff} = \nu_{oo}$) and at $I = I_r$ (unstable fixed points, $\sqrt{2I_r} = |\delta\nu|/(3D)$, where $\nu_{eff} = k/3$). One might try to evaluate $\partial\nu/\partial I$ near unstable fixed points since the stochastic layer is there. Unfortunately, this quantity diverges near unstable fixed points as shown in Fig. 8. I believe it is significant that, although they compare the theoretical predictions with numerical simulation results for three other examples, nothing further is mentioned in the report on this third model.

Fig. 8



Essentially the same argument on the resonance width is given by Keil in his report on beam-beam interaction (the second of ref. 13).* In the equation of motion for phase, he retains the average (shear) term only saying it is bigger than any other (phase-dependent) terms. For those who care to look up his report, the following "translation" of symbols would be useful to compare his formulas with ours. Symbols on the left are Keil's and those on the right are ours:

$$p \to 3, \quad q \to 0, \quad r \to -k, \quad \phi \to b, \quad \chi \to \phi$$

$$\chi' \to d\phi/d\theta \ (\equiv p/M), \quad |B| \to 36 r_o D (2I_r)^{3/2}$$

His full-width of resonance δQ_x,

$$\delta Q_x = (8\sqrt{2}/\pi|p|)\cdot|B|^{1/2} \tag{23}$$

contains an extra factor $(2/\pi)$ compared to our $2\cdot(\Delta\nu)_r$, Eq. (17). This came about as Keil demanded that one should take the maximum height of the rectangle with the same area as the stationary RF bucket. The factor is of course not essential for the present discussion. After deriving Eq. (23), Keil makes the following statement which may or may not have received Chirikov's blessing:

> "This method of calculating the width of non-linear resonances is applicable when (the shear terms) are not small. .... A method for calculating the resonance width in (cases when the shear terms are small) is shown in the Appendix."

(Here I have rewritten the parts within brackets to make the statement understandable without introducing a lengthy explanation of his notations.) This is certainly a bold jump (to me) since he is trying to extend the concept of resonance width to cases for which the under-

* One cannot help noticing a certain sloppiness in his formalism. Two different quantities, σ and τ, are used as the independent variable in two transverse directions, respectively. On p. 3, he says "Here we have assumed that all the primes mean the same thing, or alternatively that $d\sigma/d\tau \simeq 1$. This implies that we have removed the strong focusing wiggle from the calculation at this point." It is of course possible to do it exactly. For example, see Appendix A of ref. 8.

lying ideas of Chirikov may not be valid. In the following, I will try to reproduce his arguments given in the Appendix but using our notations.

From the Hamiltonian, Eq. (8), the equation of motion for the action variable I is

$$dI/d\theta = -\,\partial H/\partial b = 3D(2I)^{3/2}\sin(\phi) \tag{24}$$

where $\phi \equiv 3b - k\theta$ (see p. 19). Note that for small D and small I, the action variable is almost a constant. The equation for phase is

$$db/d\theta = \partial H/\partial I = \nu_{oo} + (4r_o)(2I) + 3D\sqrt{2I}\cos(\phi)$$

so that
$$d\phi/d\theta = 3[\nu_{oo} + (4r_o)(2I)] - k + 9D\sqrt{2I}\cos(\phi) \tag{25}$$

He then identifies the quantity within the square bracket as the shifted tune ν^*. This is reasonable since he is considering cases for which I is almost constant. Now comes the most crucial statement on the definition of the resonance width:

> "We define the width of the resonance, $2(\Delta\nu)_r$, as the range of values of ν^* over which <u>the particle motion will get locked onto it</u>. It is given by the range of ν^* which is compatible with $d\phi/d\theta = 0$ allowing that $\cos(\phi)$ varies in the range between -1 and +1."

The underline is mine and I have again changed some notations. The phrase "get locked onto it" has often been used by other people as well but, to me, it is not entirely free of ambiguities. However, the condition $d\phi/d\theta = 0$ for $-1 \leq \cos(\phi) \leq +1$ is quite clear. Two extreme values of ν^* are

$$\cos(\phi) = -1, \qquad \nu_1^* = k/3 + 3D\sqrt{2I}\ ,$$
$$\cos(\phi) = +1, \qquad \nu_2^* = k/3 - 3D\sqrt{2I}$$

so that
$$2(\Delta\nu)_r \equiv |\nu_1^* - \nu_2^*| = 6D\cdot\sqrt{2I} \tag{26}$$

This definition of the resonance width does not contain the parameter r_o of shear term and $r_o = 0$ is quite acceptable. It is important to state here that Keil intended this definition to be used only for I << beam emittance and only for lowest-order resonances since, for other cases, the phase-independent term is much larger than phase-dependent terms so that there is no need to depart from the original definition of width by Chirikov. I am sure he would strongly object if someone used the definition (26) for $I \simeq$ beam emittance. With a full understanding of this, suppose we went ahead and did that,

$$2I = \text{beam emittance}/\pi$$

and, furthermore, equated this with the unstable fixed points,

$$\sqrt{2I} \equiv \sqrt{2I_u} \simeq |\delta\nu|/(3D)$$

which is true for $D^2 >> r_o|\delta\nu|$, we would get, from Eq. (26),

$$(\Delta\nu)_r \simeq |\delta\nu|.$$

But $|\delta\nu| \equiv |\nu_{oo} - k/3|$ here is precisely the resonance width defined in the previous section (beam emittance = πR_1^2 , see p. 3), that is, the one we have already discarded as meaningless as far as the overlap criterion is concerned! (Remember my ambiguous attitude which is apparent on p. 16, "Still, I must confess ....") The central question remains: What should one use as the resonance width in the Chirikov overlap criterion for $I \simeq$ beam emittance if the shear term is very small or absent?

The mathematical tour de force shown in the report by Ruggiero,[8] entitled "Two-Dimensional Resonance Effects due to a Localized Bi-Gaussian Charge Distribution", does not really encourage one to try to digest the entire work. If one skimmed over formulas and concentrated on the major conclusions, one would be jolted by a statement regarding the sum of all resonance widths (p. 25):

> "It is not difficult to see, by inspecting (41), that all the contributions to S come only from the one-dimensional resonances, ... No explanation is offered, at the moment,

> why the bi-dimensional resonances ($p \neq 0$ and $q \neq 0$) do not contribute to the sum S."

Nevertheless, I hope I understand his work enough to explain how he defines the resonance width.

After deriving a suitable expression for the two-dimensional beam-beam field, he writes down three conditions for "fixed lines". This has already been mentioned in the previous section, p. 7. With action-angle variables $(I_x, \psi_x; I_y, \psi_y)$, two are for action variables,

$$dI_x/d\theta = 0 \quad \text{and} \quad dI_y/d\theta = 0$$

and the third is for the phase $X \equiv p\psi_x + q\psi_y - r\theta$,

$$dX/d\theta = 0.$$

The isolated resonance under consideration is of course $p\nu_x + q\nu_y = r$. Two conditions for action variables are both satisfied by taking $X = 0$ or π as usual. Two fixed lines, one with $X = 0$ in $dX/d\theta = 0$ and the other with $X = \pi$, can be expressed in the form

$$\varepsilon_{pq}^{(U)} = f^{(U)}(I_x, I_y),$$

$$\varepsilon_{pq}^{(S)} = f^{(S)}(I_x, I_y)$$

where $\varepsilon_{pq} \equiv p\nu_x + q\nu_y - r$ and two superscripts (S) and (U) are used to designate what Ruggiero calls <u>stable</u> and <u>unstable</u> fixed lines, respectively. The crucial statement on the definition of the width comes on p. 10:

> "Let us consider the three-dimensional space with ε_{pq} on the abscissa and I_x, I_y on the other two axes. Eqs. (32) and (33) represent two surfaces in this space. For fixed I_x and I_y we can go from one surface to another moving parallel to the ε_{pq}-axis. The distance $\Delta\varepsilon_{pq}$ which separates the two surfaces defines the "width" of the resonance at amplitude I_x and I_y."

The most important consequence of this definition is that the width is totally independent of the shear term regardless of its strength. This is obvious since the shear term is unchanged by the value (0 or π) of phase X. Ruggiero of course realizes that such a term exists in the beam-beam interaction. On p. 14 of his report, he states:

> "A resonance is described also by another parameter, the nonlinear tune shift. This is the distance of the center of the resonance from the linear tune. It is obtained by taking the arithmetic average of (32) and (33)."

However, this quantity does not play any role in his use of the overlap criterion. Let us apply his definition of width to our example of uncoupled resonance $3\nu = k$. Two conditions for fixed points are, from Eqs. (24) and (25),

$$\sin(\phi) = 0, \tag{24'}$$

$$(3\nu_{oo} - k) + 3\cdot(4r_o)(2I) + 9D\sqrt{2I}\cos(\phi) = 0. \tag{25'}$$

Since we are assuming $\nu_{oo} < k/3$, $\phi=0$ gives unstable fixed points and $\phi=\pi$ stable fixed points,

$$3\cdot\delta\nu^{(U)} = -3\cdot(4r_o)(2I) - 9D\sqrt{2I}\ ,$$

$$3\cdot\delta\nu^{(S)} = -3\cdot(4r_o)(2I) + 9D\sqrt{2I}\ .$$

The difference of two quantities, Ruggiero's $\Delta\varepsilon_{pq}$, is

$$|3\cdot\delta\nu^{(U)} - 3\cdot\delta\nu^{(S)}| = 18\cdot D\sqrt{2I}$$

In evaluating the sum of resonance widths, he of course divides this by p (= 3 here),*

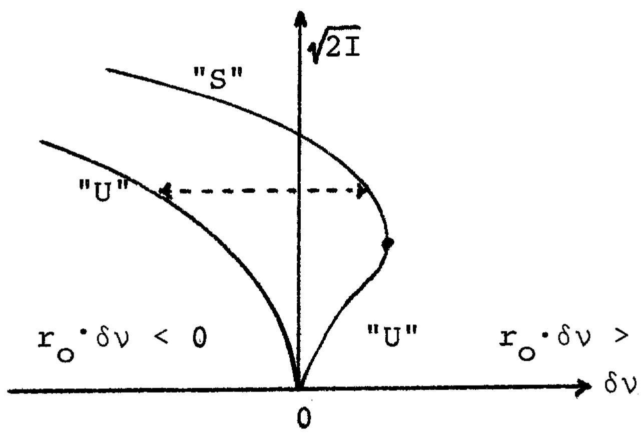
$$\text{resonance width} = 6\cdot D\sqrt{2I}$$

which is identical to Eq. (26), Keil's definition of width for cases

* See p. 23 of ref. 8.

in which the shear term is either non-existent or much smaller than phase-dependent terms. The difference is significant enough to be stated again: Keil uses the definition (26) only when he cannot use the original Chirikov definition whereas Ruggiero uses it for all cases regardless of the strength of the shear term. I have not checked if this difference explains why, in Ruggiero's calculation, there is no contribution from coupled resonances to the sum of resonance widths. However, it is not at all surprising that results obtained by Ruggiero are different from Keil's even qualitatively. Fig. 9 is a schematic presentation of Ruggiero's definition. The dotted horizontal line is an example of his resonance full-width.

Fig. 9



The article by Chao[1] from which I have already quoted a passage, is largely a review of various works on the beam-beam interaction and no novel definition of the resonance width is offered. It is interesting to see that he differentiates two types of width, one "in tune unit" and the other "in unit of $\alpha^{1/2}$" where α is a quantity proportional to our I. The former is identical to Ruggiero's definition and therefore independent of the shear term regardless of its strength. He seems to have a different physical picture in mind than Ruggiero's when he defines the "width in tune unit". After writing down the equation of motion for phase which is equivalent to our Eq. (25), he simply regards the shear term as the tune shift and the coefficient of phase-dependent term as the resonance half-width,

> "The effective tunes of a group of particles with a given amplitude α and arbitrary phase ψ occupy a spread within $\pm(\Delta\nu)_r$ around the detuned value ν^*."

He then introduces "the width in unit of $\alpha^{1/2}$" which seems to be the original Chirikov width. Although he says "Sometimes it is more convenient to define a resonance width $\delta\alpha^{1/2}$ (in unit of $\alpha^{1/2}$) to be the difference between ......", he does not tell us which definition should be used for what purpose. As a matter of fact, he does not even mention the overlap criterion.

IV. Epilogue *

This report has turned out to be a compilation of troubled monologues. I _am_ troubled by deft touches of verisimilitude which appear from time to time in works dealing with nonlinear resonances, especially when they are not accompanied by solid substance. There could be many different conclusions drawn from this study by different readers. There is certainly a question of who is "right" and who is "wrong" but I cannot provide an answer to that question. In writing this report, it was never my intention to criticize what others have done. After all, I have nothing better to offer in place of their works either in the definition or in the usage of resonance width. Many people including Chirikov agree that the overlap criterion is simply an order-of-magnitude estimate and no more. At the same time, the fact that it is only an order-of-magnitude formula should not be regarded as a license for its careless treatment.

* "a concluding section that rounds out the design of a literary work" - Webster's Dictionary -

References

1. A. W. Chao, Nonlinear Dynamics and the Beam-Beam Interaction, edited by M. Month and J. C. Herrera, AIP Conference Proceedings No. 57, American Institute of Physics, 1979, p. 42.
2. P. A. Sturrock, Annals of Physics, 3 (1958), 113.
3. W. P. Lysenko, Particle Accelerators, 5 (1973), 1.
4. S. Ohnuma, Fermilab TM-507, July 15, 1974.
5. S. Ohnuma, Fermilab TM-910, October 15, 1979.
6. S. Ohnuma, Fermilab FN-329, August 18, 1980.
7. T. L. Collins and D. A. Edwards, Fermilab TM-614, September 1975.
8. A. G. Ruggiero, Fermilab FN-258, April 24, 1974.
9. G. Guignard, CERN-ISR-MA/75-35, 31st August, 1975; CERN 78-11, 10 November, 1978.
10. B. V. Chirikov, Physics Reports, 52 (1979), 263.
11. S. Ohnuma, Fermilab UPC No. 117, December 10, 1979.
12. B. V. Chirikov, E. Keil, and A. M. Sessler, ISR-TH/69-59, October 15, 1969.
13. E. Keil, CERN/ISR-TH/72-7, 7th February, 1972; CERN/ISR-TH/72-25, July, 1972.
14. L. C. Teng, IEEE Trans. Nucl. Sci., NS-20 (1973), 843; Fermilab UPC No. 67, October 20, 1978.
15. E. D. Courant and H. S. Snyder, Annals of Physics, 3 (1958), 1; S. Ohnuma, Fermilab TM-749, October 24, 1977.
16. J. M. Greene, PPPL-1489 (Plasma Physics Laboratory, Princeton University), November 1978. This has been published in Jour. Math. Phys., 20 (1979), 1183.